基于 ResNet 和 SeNet 的图像分类研究

代英 [1]

[1]（北京化工大学 北京 100029）

摘要

图像分类和识别在现代社会中具有重要意义。已经有许多优秀的卷积神经网络工作来优化图像分类的准确性，其中一位杰出的代表是 ResNet[1]，它大幅增加了神经网络的深度，从而极大地提高了神经网络的性能。与此同时，还有一些可插拔的性能优化子模块可以帮助优化所有网络，其中一个杰出的代表是 SeNet[3]。然而，在面对现实世界中的复杂场景时，它们并不总是表现良好。本文的主要工作是研究如何有效提高卷积神经网络（ResNet）在一些特殊场景（小图片、高噪声图片）中的识别性能，并尝试分析一些神经网络的底层机制。

关键词：图像分类 卷积神经网络

分类号：

Image classification research based on ResNet and SeNet

Dai ying[1]

[1](Beijing University of Chemical Technology, Beijing 100029, China)

Abstarct

Image classification and recognition are of great significance in modern society. There have been many excellent convolutional neural network works to optimize the accuracy of image classification, one of the outstanding representatives is ResNet[1], which greatly increases the depth of the neural network, thereby greatly improving the performance of the neural network. At the same time, there are some pluggable performance optimization sub-modules that can help optimize all networks, one of the outstanding representatives is SeNet[3]. However, they do not always perform well when faced with complex scenarios in the real world. The main work of this article is to study how to effectively improve the recognition performance of convolutional neural networks (ResNet) in some special scenes (small pictures, high-noise pictures), and try to analyze the underlying mechanisms of some neural networks.

Keywords: Image classification, convolutional neural networks

1.引言

在深度学习中，卷积神经网络（Convolutional Neural Network，CNN）是最常用的模型之一，已经在计算机视觉和语音识别等领域取得了极大的成功。本文将简要介绍一些经典和最新的 CNN 模型，从 AlexNet 到 ResNet 和 SeNet。

AlexNet[5]是 CNN 领域的里程碑，在 2012 年的 ImageNet 大规模视觉识别挑战赛中获得了第一名。随后，基于 AlexNet 的许多改进被提出，例如 VGG[6]、GoogLeNet[7]和 ResNet[1]。特别是 ResNet 是一种更深层次的 CNN，通过使用残差连接来解决梯度消失的问题，在 ImageNet 比赛中取得了出色的成绩。ResNet 已经成为许多最先进的 CNN 模型的基础构建块，展示了深度学习在计算机视觉中的重要性。

SENet[3]，或称为 Squeeze-and-Excitation Network，是一种最近的卷积神

经网络架构，在 ImageNet 分类任务上取得了最突出的效果。SENet 引入了一个名为 Squeeze-and-Excitation 块的新模块，通过明确建模通道之间的相互依赖关系，自适应地重新校准通道级特征响应。这种技术允许网络在抑制不太有用的通道的同时，赋予信息丰富的通道更重要的权重，从而提高准确性和效率。SENet 已成为现代 CNN 的另一个重要构建模块，展示了深度学习技术的持续演进。除了这些模型之外，还有许多其他的 CNN 架构，如 DenseNet[4]、MobileNet[2]和 EfficientNet[8]，它们在不同的应用场景中表现出色。然而，对于复杂且多变的实际场景，这些优秀的算法仍然无法在所有场景中表现出色。本文的主要目的是通过对某些特定场景（小图片、大噪声）进行网络调整，尝试分析这些场景中的一些通用方法和相关网络的底层机制。

2.相关工作

2.1ResNet

ResNet 是一种深度神经网络架构，于 2015 年提出。ResNet 的全称是“残差网络”，它旨在通过引入残差块来解决深度神经网络中的梯度消失问题。在传统的神经网络中，每个层都对输入进行转换并输出新的特征表示。当网络变得非常深时，这些转换会导致输入信号逐渐消失，从而产生梯度消失问题。为了解决这个问题，ResNet 引入了残差块。在 ResNet 中，每个残差块包含两个分支：主分支和跨层连接分支。主分支对输入执行一系列变换，并将结果添加到跨层连接分支的输出，以最终获得残差块的输出。这种设计使得网络更容易学习恒等映射，即输入和输出相等的情况。如果残差块中没有发生变化，则可以通过跨层连接将输入直接传递到输出，从而避免信息丢失和梯度消失问题。这种设计使 ResNet 能够训练非常深的神经网络，并在几个计算机视觉任务中表现良好，例如图像分类，目标检测和语义分割。

2.2 SeNet

在大多数利用卷积神经网络（CNN）处理图像的研究中存在一个问题：忽视了不同通道之间的相互关系。为了解决这个问题，SeNet 提出了 Squeeze-and-Excitation（SE）模块，通过学习每个通道的权重加强了通道之间的相互关系，提高了模型的表达能力。SE 模块包括两个步骤：压缩（squeeze）和激励（excitation）。压缩操作通过全局平均池化将每个通道的特征图压缩为一个单一的值，得到每个通道的权重。激励操作使用多层感知机（MLP）来结合每个通道的特征图，并根据每个通道的权重进行加权。具体而言，压缩操作首先对输入特征图的每个通道进行全局平均池化，然后通过全连接层和 ReLU 激活函数处理每个通道的结果值，得到每个通道的权重。激励操作使用 MLP 对每个通道的特征图进行加权和组合，得到增强了通道之间相互关系的输出特征图。MLP 的输入是从压缩步骤中获得的每个通道的权重，输出是与输入特征图相同大小的权重向量。通过将 SE 模块嵌入到其他网络结构中，可以使用 SE 模块。具体而言，SE 模块可以插入到卷积神经网络的每个模块中，使模型在学习特征表示的同时自适应地学习通道之间的关系。

图 1 是嵌入 SE 模块的典型 SeNet 结构。SE 模块已被证明在各种计算机视觉

任务中表现良好，包括图像分类，对象检测和语义分割。在 ImageNet 图像分类挑战中，SE 模块将 top-1 和 top-5 的准确率分别提高了约 2.5% 和 1.0%。在 ImageNet 图像分类挑战中，SE 模块将 top-1 和 top-5 的准确率分别提高了约 2.5% 和 1.0%。此外，SE 模块可以应用于各种深度学习模型，如 MobileNet 和

DenseNet。

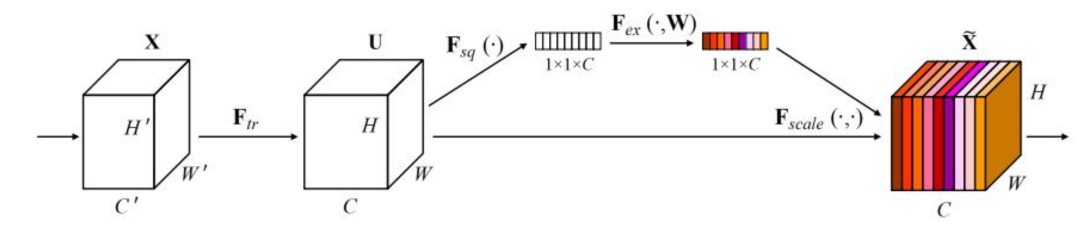


图 1：SeNet 结构

3. 数据集

CIFAR-10 数据集是一组常用于训练机器学习和计算机视觉算法的图像集合。它是机器学习研究中最广泛使用的数据集之一[1][2]。CIFAR-10 数据集包含了 60,000 张 32x32 像素的彩色图像，分为 10 个不同的类别[3]。这 10 个不同的类别分别代表飞机、汽车、鸟类、猫、鹿、狗、青蛙、马、船和卡车。每个类别有 6,000 张图像[4]。计算机算法在识别照片中的物体时通常通过示例学习。CIFAR-10 是一组可以用来教导计算机如何识别物体的图像。由于 CIFAR-10 中的图像分辨率较低（32x32），这个数据集可以让研究人员快速尝试不同的算法，以查看哪种算法效果最好。CIFAR-10 是 80 Million Tiny Images 数据集的一个带有标签的子集。在创建该数据集时，学生们被要求对所有图像进行标注[5]。各种类型的卷积神经网络通常在识别 CIFAR-10 中的图像方面表现最好。该数据集被划分为五个训练批次和一个测试批次，每个批次包含 10000 张图像。测试批次中包含每个类别随机选择的 1000 张图像。训练批次以随机顺序包含剩余的图像，但某些训练批次中可能包含来自某个类别的图像比其他类别多。在这些训练批次中，每个类别恰好包含 5000 张图像。图 2 是数据集中的类，以及每个类中的 10 个随机图像：

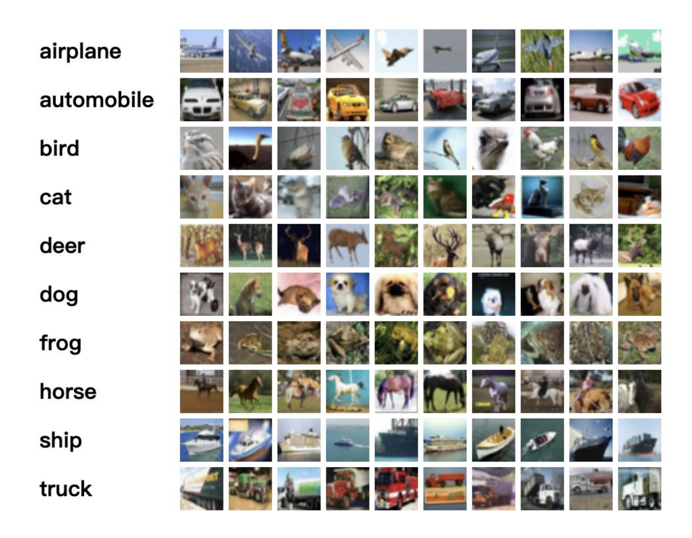


图 2：CIFAR-10 数据集

4. ResNet 的效果

在本节中，基于 PyTorch 实现了 ResNet。研究了对 ResNet 的某些参数和结构的影响，并对其结构和参数进行了修改。将实现的 ResNet 和基准 ResNet 在 CIFAR-10 数据集上进行了评估。最后分析了这些改动对模型的分类性能的影响。

4.1 ResNet 的结构

ResNet 的结构可以分为两种类型：基本块（Basic Block）基于两个卷积层，瓶颈块（Bottleneck Block）基于三个卷积层。基本块适用于 ResNet18 和 ResNet34 等浅层网络，而瓶颈块适用于 ResNet50 和 ResNet101 等深层网络。

基本块和瓶颈块的结构如图 3 所示，左边是基本块，右边是瓶颈块。基本块由两个由 3x3 卷积层组成的残差模块构成。输入和输出通道数相同。如果输入和输出的尺寸不一致，应该在输入中添加一个 1x1 卷积层来匹配尺寸。

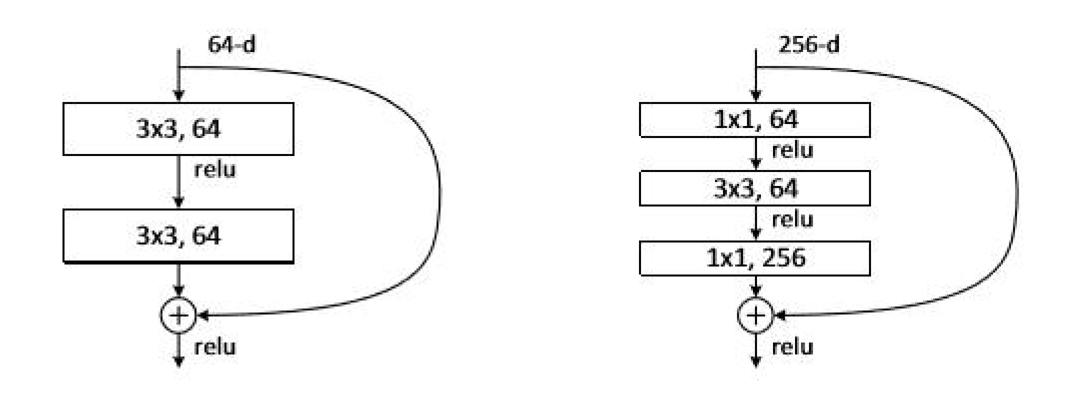


图 3：基本块和瓶颈块的结构

与基本块不同，瓶颈块在每个残差连接中添加了一个 1x1 卷积层来进行维度匹配。此外，瓶颈块使用 1x1 + 3x3 + 1x1 的结构来替代两个 3x3 卷积层，从而降低了计算复杂度和参数数量，并增加了网络的非线性。瓶颈块的缺点是，在浅层网络中使用 1x1 卷积层可能会丢失一些特征信息，并导致性能下降。

实现的 ResNet 具有以下结构：

输入层：输入图像的大小为 224x224x3，其中 3 表示 RGB 通道。

卷积层：第一层是一个 3x3 的卷积层，步长为 2，使用 64 个卷积核，填充为 3，这使得输入和输出的尺寸保持一致。该层后面跟着一个批量归一化层和 ReLU 激活函数。

池化层：接下来是一个 3x3 的最大池化层，步长为 2，填充为 1，可以将输入尺寸减半。

残差块：然后有 4 个残差块，每个块包含多个具有相同结构的残差单元。每个残差单元由 2 个卷积层和一个恒等映射组成，其中第一个卷积层的步长可以设置为 2，进一步减小特征图的大小。每个残差块中第一个残差单元的通道数为 64，并且随着残差块的深度增加，通道数翻倍，直到最后一个残差块具有 512 个通道。

全局平均池化层：有一个全局平均池化层，将最后一层的特征图转换为一个 1x1x512 的张量。

最后，有一个全连接层，将 512 维的特征映射到类别的数量上。

4.2 评估

实验使用 PyTorch 框架，训练集使用了前文中描述的 CIFAR-10 数据集，该数据集中的图像大小为 32x32 像素。总共训练和评估了两个模型：(1) 基准的 ResNet-18 模型；(2) 将我们的修改集成到 ResNet-18 中。优化器使用随机梯度下降（SGD）算法，学习率为 0.1，动量为 0.9，权重衰减为 5e-4。我们使用两个不同的学习率进行训练，以比较它们的效果。模型从头开始训练，共进行 200 个 epoch。为了确保数据被标准化，我们对训练图像进行了标准化处理。

（1）训练结果

图 4、5 展示了在训练集上的结果。训练集上的结果并不是很有意义，因为两个模型都可以达到 100%的准确率。

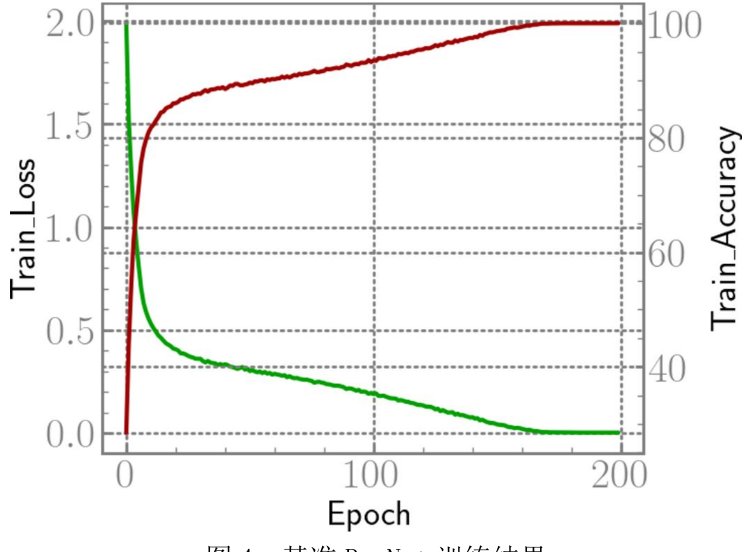


图 4：基准 ResNet 训练结果

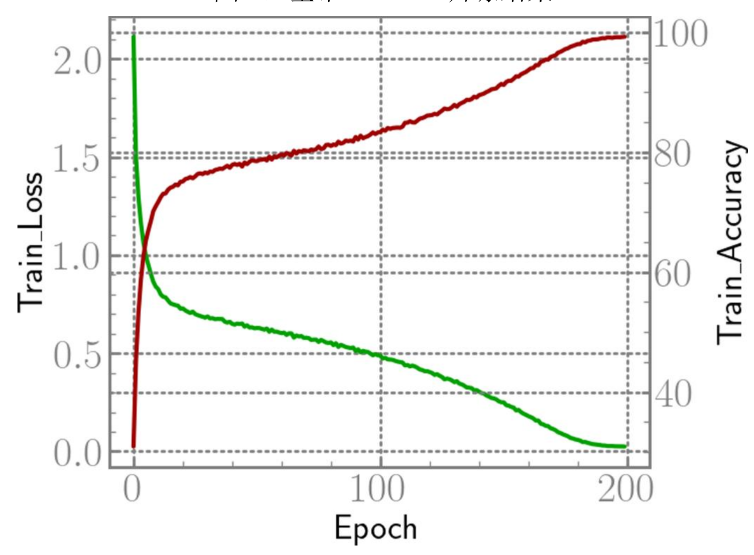


图 5：改动后 ResNet 训练结果

（2）测试结果

图 6 和图 7 展示了基准 ResNet 和我们的 ResNet 在测试集上的测试结果，图 8 展示了两个训练模型的准确率比较结果。通过比较基准 ResNet 和改动的 ResNet 在测试集上的分类结果，我们可以发现我们的 ResNet 相对于基准 ResNet 的准确率提高了 6%，同时损失率也更低。

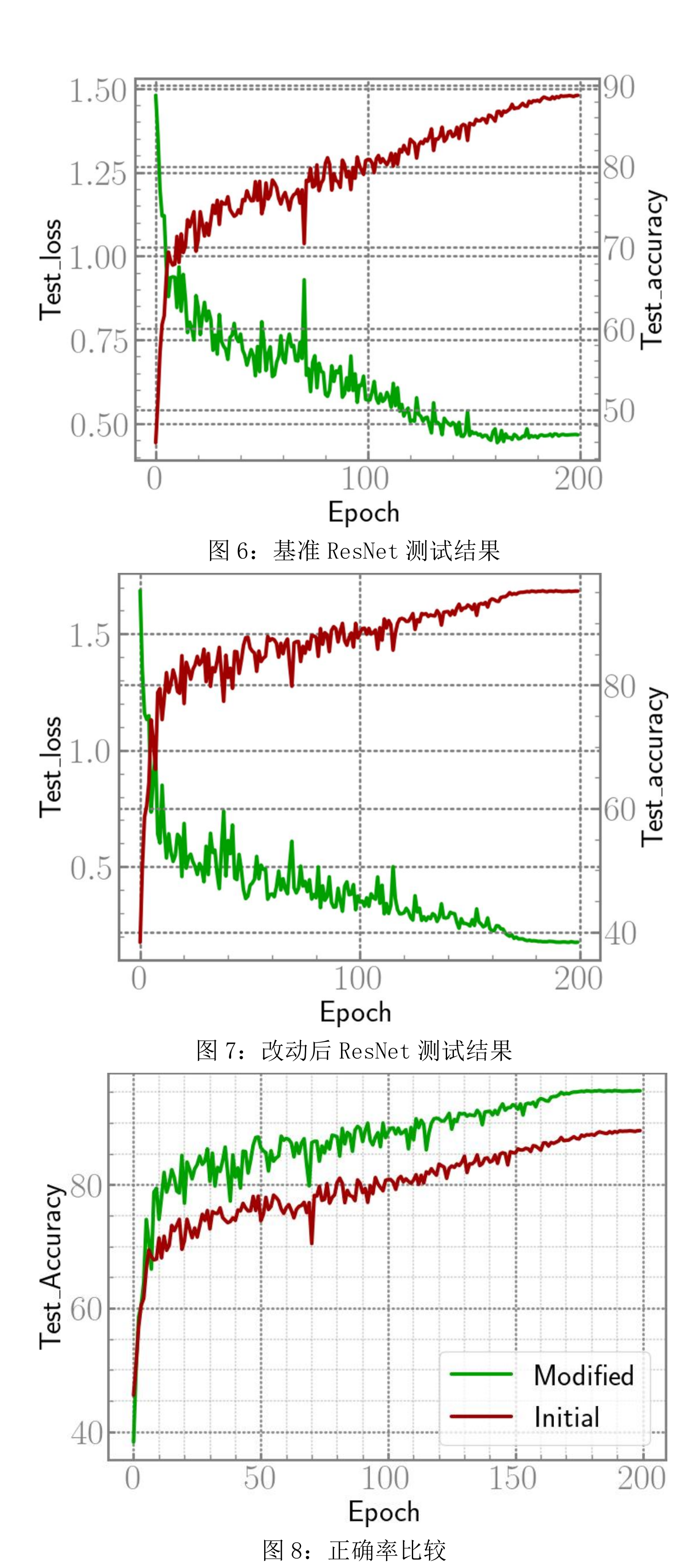


图 6：基准 ResNet 测试结果

图 7：改动后 ResNet 测试结果

图 8：正确率比较

4.3 讨论

我们的 ResNet 实现相比基准 ResNet 使用了更小的卷积核大小。改动使用了

3x3 的卷积核，而基准 ResNet 使用了 7x7 的卷积核。使用 7x7 的卷积核可以在输入层对输入图像进行下采样，减少计算量，同时保持较大的感受野，以尽可能保留原始图像的信息。而使用 3x3 的卷积核可以在输入层进行更精细的特征提取，保留边缘信息，减少模糊和不确定性。改动实现的较小的卷积核大小可能会增强对小物体的检测能力，但在检测较大物体方面可能不如基准实现表现好。

在改动的 ResNet 实现中，在输出层之前添加了一个全局平均池化层。使用池化层和全连接层作为输出层可以对特征图进行全局平均池化，减少模型参数的数量，降低过拟合的风险，并通过特征图的整体平均值捕捉更多的特征信息。此外，使用平均池化使得模型的特征表示更加通用和可移植。

将全连接层作为输出层可以直接线性转换特征图以输出预测结果，但需要更多的参数和计算量，并可能需要调整输出维度、类别数量或回归范围。这也可能导致输出层和其他层之间梯度更新不一致，需要使用残差连接来解决这个问题。

5. SeNet 的影响

5.1 将 SeNet 嵌入 ResNet

如图 9 所示，SENet 提供了四种方法将 SE 模块嵌入 ResNet 架构中，分别是 SE、SE-Pre、SE-Post 和 SE-Identity。

在 SEResNet 设计中，SE 模块被插入到 ResNet 架构的每个残差块中，位于最后一个卷积层之后和最后的加法操作之前。SE 模块的输出然后与残差连接相加，得到块的最终输出。

在 SE-Pre ResNet 设计中，SE 模块被插入到预激活 ResNet 架构的每个预激活残差块中，位于第一个批归一化层之后和最后一个卷积层之前。SE 模块的输出然后与残差块的输入相加，位于第一个批归一化层之前。

在 SE-Post ResNet 设计中，SE 模块被添加到网络中每个残差块的最后一个卷积层之后。残差块的输出然后与 SE 模块的输出相加，得到块的最终输出。这种方法与 SE ResNet 类似，只是 SE 模块添加在最后一个卷积层之后，而不是最后的加法操作之前。

在 SE-Identity ResNet 设计中，SE 模块被添加到每个残差块中的身份快捷连接。SE 模块的输出然后被添加到身份快捷方式以获得块的最终输出。该方法类似于 SEResNet，除了 SE 模块被添加到标识快捷方式而不是残差块的主分支。

在接下来的评测中，我们采用 SE ResNet 设计将 SE 集成到 ResNet 18 中，这也是 SENet 原始论文中的标准设计。

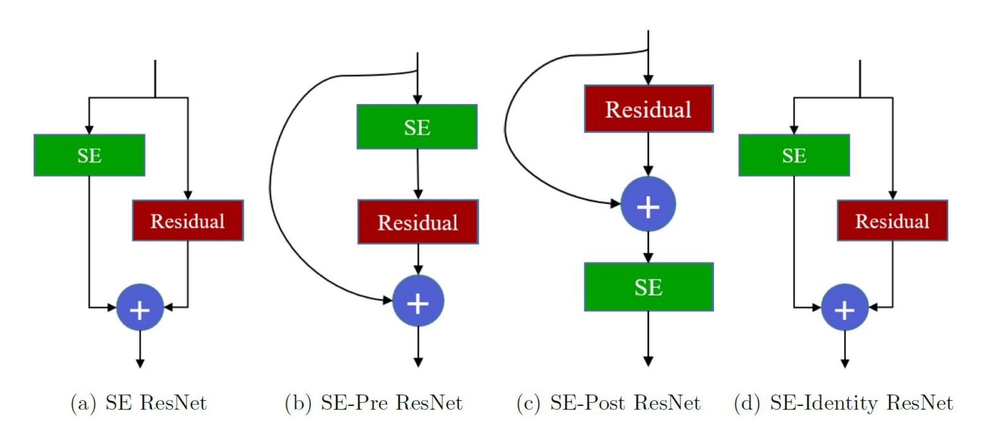


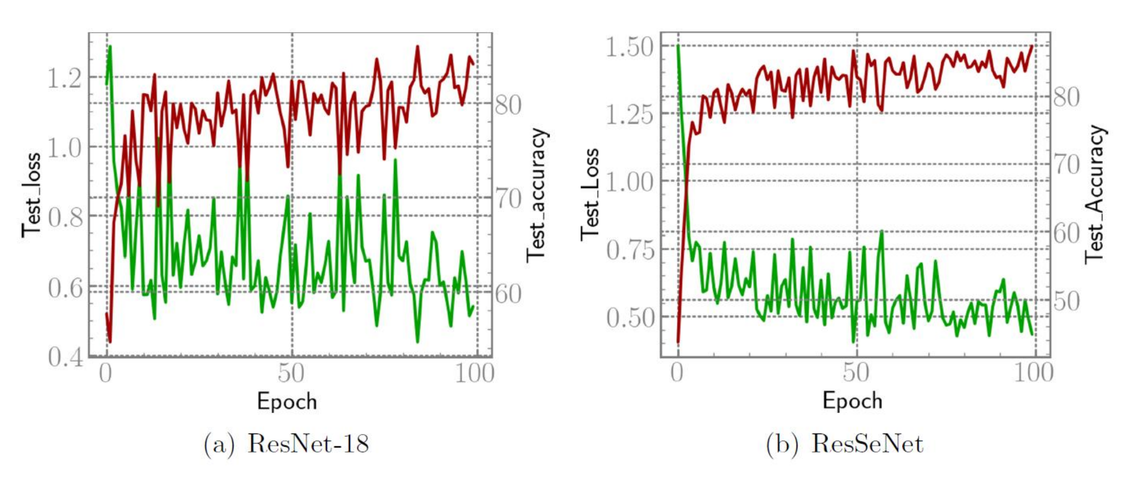
图 9 将 SE 嵌入 ResNet 的四种设计

5.2 评估

CIFAR-10 数据集中的原始图像大小为 32x32 像素，我们将其调整为 224x224 像素进行训练。放大图像会引入一些噪音到数据中，这可能会对模型的性能产生影响。然而，我们相信这有助于评估 SENet 架构在提高 ResNet-18 模型分类性能方面的有效性。

总共训练和评估了两个模型：(1) 基准的 ResNet-18 模型，(2) 将 SE 集成到 ResNet-18 模型中。优化器使用随机梯度下降（SGD）算法，学习率分别为 0.1 和 0.001，动量为 0.9，权重衰减为 5e-4。我们使用两个不同的学习率进行训练，以比较它们的效果。模型从头开始训练，共进行 100 个 epoch。为了确保数据归一化，我们对训练图像进行了标准化处理。

图 10 展示了两个模型在测试集上随着 epoch 的增加的训练准确率和损失。这两个模型都是使用学习率 lr=0.1 进行训练的。由于较大的学习率和图像放大引入的噪音干扰，ResNet 模型在训练过程中的准确率呈现出明显的波动。可以看到，在将 SENet 模块集成后，模型的波动显著减少，平均准确率有所提高。图 11 展示了使用学习率 lr=0.001 进行训练的结果。可以观察到准确率的严重波动已经消失，在经过 50 轮训练后，准确率稳定下来。此时，ResSENet 仍然比 ResNet 具有更高的准确率，提高了 1%。图 12 显示了两个模型在不同学习率下准确率的可视化比较。实验结果表明，在各种场景下，添加 SENet 是有效的。它对噪音具有显著的抑制作用，并使准确率有一定提高。



(a) ResNet-18　　(b) ResSeNet

图 10：学习率 lr=0.1 的训练结果

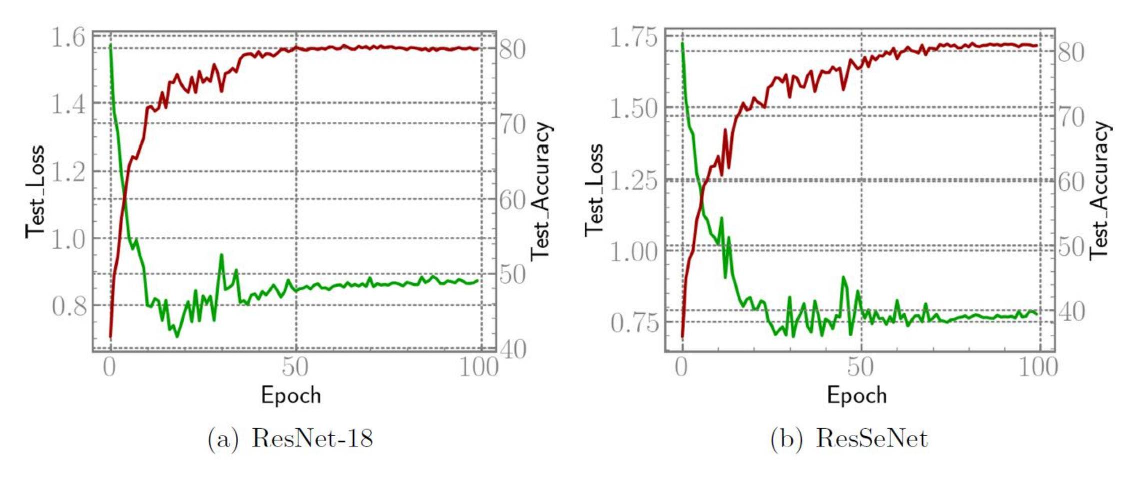


图 11：学习率 lr=0.001 的训练结果

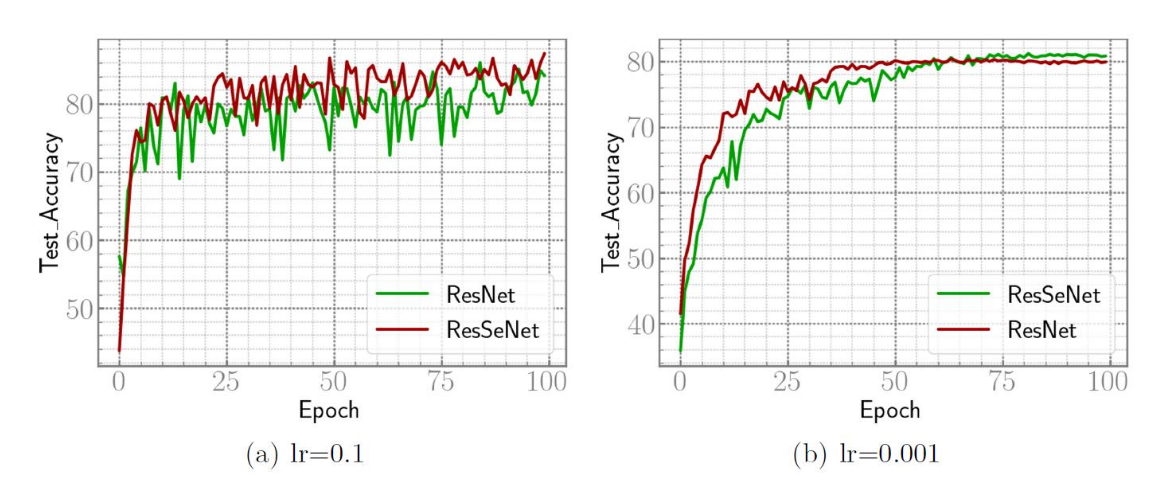


图 12：不同学习率模型正确率对比

5.3 讨论

在这项研究中，我们评估了将 SE 块添加到 ResNet 中对图像分类性能的影响，并考察了不同学习率下的效果。我们的实验结果表明，添加 SENet 显著提高了模型的性能。具体而言，我们观察到准确率的波动明显减少。

SENet 性能提升的一个主要原因是 SE 块提供的注意力机制。SE 块可以有选择地放大信息丰富的特征，同时抑制不太有用的特征。这个机制可以帮助网络关注最重要的特征，从而提高特征的可辨识性，并使网络对噪音更加鲁棒。

此外，我们的结果显示，SENet 的添加在高噪音水平的场景中特别有效，其中准确率的波动更加明显。SE 块改善了 ResNet 的整体稳定性，通过减小训练过程中准确率波动的幅度来展示。这表明 SENet 可以帮助减少噪音对模型性能的影响，这在深度学习的实际应用中是一个重要的考虑因素。

总而言之，我们的研究表明，将 SENet 添加到 ResNet 中可以显著提高图像分类任务中的性能。SE 块提供的注意力机制在增强特征的可辨识性和减少噪音对模型的影响方面起着关键作用。这些发现对于各种应用的深度学习模型设计具有重要的影响。

6. 结论

正如我们在 4.3 节中讨论的那样，较小的图像可能需要较小的卷积核和尽可能小的步长，以防止遗漏一些重要的图像特征。为了在保持广泛感受野和尽可能

保留原始图像信息的同时减少计算复杂性，可以使用 7x7 的卷积核进行下采样。另一方面，使用 3x3 的卷积核可以实现更精确的特征提取，保留边缘细节并减少模糊和歧义。虽然我们实现中使用较小的卷积核可能增强了检测较小物体的能力，但在检测较大物体方面可能不如基准方法效果好。

正如我们在 5.3 节中提到的，SENet 性能提升的一个主要原因是 SE 块提供的注意力机制，它可以有选择地放大信息丰富的特征，同时抑制不太有用的特征。这个机制帮助网络关注最重要的特征，提高特征的可辨识性，并使网络对噪音更加鲁棒。我们的结果还表明，在高噪音水平的场景中，添加 SENet 特别有效，其中准确率的波动更加明显。

此外，SE 块增强了 ResNet 的整体稳定性，通过减小训练过程中准确率的波动来证明。这表明 SENet 可以帮助减轻噪音对模型性能的影响，这对于深度学习的实际应用是一个重要的考虑因素。

总而言之，我们的研究表明，将 SENet 添加到 ResNet 中可以显著提高图像分类任务的性能。SE 块提供的注意力机制在增强特征的可辨识性和减轻噪音对模型的影响方面发挥着关键作用。这些发现对于设计各种应用的深度学习模型具有重要的意义。

参考文献：

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[2] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

[3] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018.

[4] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In Proceeding of the IEEE conference on computer vision and pattern recognition, pages 4700–4708, 2017.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6):84–90, 2017.

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[7] ChristianSzegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, DragomirAnguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolu- tions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.

[8] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, pages 6105–6114. PMLR, 2019.